

統計学教育の実践例と将来に向けた方向性

加藤 淳一

(久留米大学 商学部)

One Example on Education in Statistics and Discussion for a Future Direction

Junichi KATO

Department of Commerce, Faculty of Commerce
Kurume University

【要約】 本論文の目的は、私立大学文系学部での著者自身の統計学教育の実践例を議論の出発点として例示した上で、先行研究に依拠しつつ将来の統計学教育（とりわけ、経営意思決定に統計学を利用するという立場からの統計学教育）の方向性を探ることである。まず第 2 章で、議論の出発点として、著者自身が私立大学文系学部で行っている授業の実践例について報告をする。シラバスの各回の内容とともに各授業回の授業の構成についても触れる。

次に第 3 章で、先行研究に依拠して、この授業を批判的に検討する。第 3 章での検討は、著者自身の授業（経営統計論 I の授業）が統計学者を目指す学生ではなく、統計学を経営意思決定に利用しようとしている学生を対象としているという観点からなされる。現行の授業がこうした学生の目的と一致しているのか。一致していないとしたならば、どのような授業にすれば良いのか。これらについて検討する。こうした検討を通じて、私立大学文系学部の学生への統計学教育（とりわけ、経営意思決定に統計学を利用するという立場からの統計学教育）の手がかりを探る。

最後に第 4 章で、結論として再び目的に戻り、目的と対応づけながらここまでの議論で明らかになったことを整理する。本論文の議論により、パフォーマンス課題により学生の経営意思決定に統計学を活用する能力をパフォーマンスという形で可視化して、そのパフォーマンスに対してルーブリックという評価基準により評価する。このような授業を今後の授業の一つの手がかりとして示した。そして、マーケティング研究者が、本論文のような教育に関連した論文を執筆した理由を手短に述べて締めくくる。

【キーワード】 データサイエンス, 統計学教育, 経営統計学, ルーブリック,
パフォーマンス評価

I. 目的

本論文の目的は、私立大学文系学部での著者自身の統計学教育の実践例を議論の出発点として例示した上で、先行研究に依拠しつつ将来の統計学教育（とりわけ、経

営意思決定に統計学を利用するという立場からの統計学教育)の方向性を探ることである。

我が国の大学教育に限っても、滋賀大学データサイエンス学部¹の新設をはじめとして、データサイエンスに注目が集まっている。だが、データサイエンス学部あるいはその学問分野に注目するだけでなく、大学においてデータサイエンスをいかに学修(あるいは教育)するのかを考えなければならない。

ただし、データサイエンスの学修(あるいは教育)を一編の論文で扱うのは無理がある。その理由として、ここでは次の2点を指摘する。1点目は、データサイエンスは広範で多様な専門領域の複合体であり、その専門領域を一編の論文では網羅できない。2点目に、大学学部教育と一口に言ってみても、その学生の知識は多様である。とりわけ、私立大学文系学部の学生は、他の学生たちと異なる一つの制約条件を持つ。それは、大学入試の段階で数学を必要とせずに入學した学生が一部に見られることである。

これら2点を考慮に入れ、本論文はまず対象領域としてデータサイエンスの広範で多様な専門領域の中でも中核の一つの統計学の学修(あるいは教育)に焦点を絞る。次に、考察の対象を私立大学文系学部の学生に限定する。統計学に絞る理由は、データサイエンスにとって中核であるのに加えて、既に広く大学で学修(あるいは教育)されているからである。また、私立大学文系学部学生を対象とする理由は、私立大学文系学部学生に実行可能な学修(あるいは教育)ならば、他の学部学生へ応用可能だと思えるからである。よって、本論文は、データサイエンスの学修(あるいは教育)についての考察の第一歩として、私立大学文系学部での統計学の学修(あるいは教育)について考察を試みる。

私立大学文系学部の学部学生を対象とする場合、高等学校で学ぶはずの数学の知識は一切与件にできない。したがって、選択肢は大きく2つとなる。1つ目は、統計学を学ぶ上で必要な数学の知識を身につけた上で統計学の学修(あるいは教育)を行う。2つ目は、数学の知識を必要としない範囲で統計学を学修(あるいは教育)する。

これらのうちどちらかを選択しなければならない。こうした問題への著者なりの工夫を含めて、本論文の出発点として著者自身の大学で行ってきた授業を実践例として例示するところから取りかかる。その上で、先行研究に依拠しつつ将来の統計学教育(とりわけ、経営意思決定に統計学を利用するという立場からの統計学教育)の方向性を探る。

改めて本論文は次の構成で論を進める。まず第2章で、議論の出発点として、著者自身が私立大学文系学部で行っている授業の実践例について報告をする。シラバスの各回の内容とともに各授業回の授業の構成についても触れる。次に第3章で、先行研究に依拠して、この授業を批判的に検討する。第3章での検討は、著者自身の授業(経営統計論Ⅰの授業)が統計学者を目指す学生ではなく、統計学を経営意思決定に利用しようとしている学生を対象としているという観点からなされる。現行の授業がこうした学生の目的と一致しているのか。一致していないとしたならば、どのような授業にすれば良いのか。これらについて検討する。こうした検討を通じて、私立大学文系学部の学生への統計学教育(とりわけ、経営意思決定に統計学を利用すると

いう立場からの統計学教育)の手がかりを探る。最後に第4章で、結論として再び目的に戻り、目的と対応づけながらここまでの議論で明らかになったことを整理する。そして、マーケティング研究者が、本論文のような教育に関連した論文を執筆した理由を手短に述べて締めくくる。

II. 実践例

本論文の議論の出発点として、ここでは著者自身が担当している経営統計論Ⅰの講義を取り上げる。まず1節で、経営統計論Ⅰの授業目的に始まり、概要、そして到達目標を確認する。次に2節と3節で、15回(半期分)の各回で取り扱う授業内容を簡単に説明する。最後に4節で、各回の授業構成を説明する。これらの現行の授業の説明が、後の章でとりわけ経営意思決定に統計学を用いるという観点から、現行の授業の問題点とその問題点を解決する手がかりを吟味していくときの出発点となる。

1. 授業「経営統計論Ⅰ」の目的

この経営統計論Ⅰの目的は、加藤(2017, 151 頁)によると「経営統計は経営意思決定において重要な位置を占めている。経営統計は、客観的な指標に基づいた意思決定の基礎であり、その理解はビジネス・リーダーとして必要な知識となっている」としている。ここで「ビジネス・リーダー」という言葉を用いているのは久留米大学商学部の教育理念「新しい社会をひらくビジネス・リーダーの育成」と整合的である。つまり、この科目が久留米大学商学部の教育理念とむすびつきをもって開講されていることを示している。

そして、概要は加藤(2017, 151 頁)によると「本授業では、統計について扱ったテキ

表1 経営統計論Ⅰの授業計画(出典、加藤(2017, 151 頁)のシラバスの一部分を使用)

授 業 計 画	1 講義の概要： 半期の授業の大まかな内容、各回の授業の進め方、そして成績評価方法について理解する
	2 統計学とは何か： 統計学とはどのような学問なのかを理解する
	3 統計データの分類： 質的データと量的データの違い、4つの尺度、統計データと変数と観測値の定義と表記法を理解する
	4 統計データの集計： 質的データと量的データの集計方法を理解する
	5 統計表のグラフ表現： 構成比グラフ、相対度数、累積相対度数、度数分布、ヒストグラムの作成方法を理解する
	6 中心位置の統計量： 平均値と中央値の定義、平均値の性質を理解する
	7 変化を表す統計量： 変化を表す統計量(変化率、寄与度、そして寄与率)について理解する
	8 中間テスト： ここまでの学修内容の理解度を確認する
	9 散らばりの統計量： 散らばり、分散、標準偏差について理解する
	10 標準偏差の活用1： チェビシェフの不等式、変動係数、標準化について理解する
	11 標準偏差の活用2： チェビシェフの不等式、変動係数、標準化について理解する
	12 散らばりのグラフ表現： 四分位範囲と箱ひげ図の作成方法を理解する
	13 2変数の関連性1： 相関係数の計算方法とその性質を理解する
	14 2変数の関連性2： 相関係数の計算方法とその性質を理解する
	15 統計データの整理： 統計データの整理方法を理解する

ストを基礎にしつつ、記述統計学とは何かを理解する」としている。なお、テキストは稲葉(2012)である。更に、到達目標は加藤(2017, 151 頁)によると「本授業の目標は、客観的な根拠に基づいた経営意思決定の基礎として記述統計学の考え方を理解することである」としている。つまり、授業全体で目指している到達点は、記述統計学の基本的な知識の習得である。

この目的や目標を達成するための各 1 回分の授業構成については、後ほど節を改めて説明する。経営統計論 I のシラバスの中の授業計画は表 1 のようである。授業計画の詳細は、表 1 をご覧いただければ容易に理解できるところであろう。

2. 授業「経営統計論 I」の 15 回（半期）のうち前半 8 回の構成

ここでは引き続いて、15 回の授業で取り扱う大まかな内容を確認する。基本的に、使用しているテキスト（稲葉, 2012）の目次に沿う内容となるが、今後の議論の出発点となるので 1 回目から確認する。1 回目は、経営統計論 I の半期の授業で学修する概要、各 1 回分の授業をどのようにすすめるのか、そして最終的に成績をどのように評価するのか評価方法について説明し理解させている。1 回目の授業を通じて、受講生が全体像を理解できた上で、各回の授業を受講できるようにする。

2 回目は、広く大学の統計学の授業で取り扱われている推測統計学と記述統計学の違いや本授業（経営統計論 I）で扱う記述統計学の特徴を説明し理解させている。そのうえで、今後の授業の中で使用する数学記号について説明し、次回までの宿題として暗記させる。これら記号は、必要になるたびごとに一覧表で確認しても差し支えない。だが、経験的に徐々に一覧表に戻り確認するのが億劫になり、結局分からなくとも戻り見なくなる。あるいは、記号がすぐに分からないことにより、教員の説明がすんなり理解できず、「わからない」「難しい」という誤った感覚を生み出す源泉となっている。このような認識がある。そこで、これからの説明を理解していく前提として、必要最低限度の記号を覚えてくるように指導している。

3 回目は、統計データの分類を説明し理解させている。記述統計学の場合、まず手元のデータがどのようなデータなのかを理解しなければ、以後の分析はできない。分類はおおまかに質的データと量的データの 2 分類を説明し、次に質的データを名義尺度と順序尺度に、量的データを間隔尺度と比率尺度に分類して説明する。これらの分類を受講生自身でできるように説明し、理解させている。加えて、データの種類として、時系列データやクロスセクショナルデータ、パネルデータの別を説明する。

4 回目は、質的データと量的データの集計方法について説明している。質的データの集計方法として、観測値の頻度（度数）を統計表として集計する方法、統計表の各部分の名称（例えば、表題、表側、表頭など）の説明を行う。これに対して、量的データの集計方法として、階級を区分して統計表として集計する手順とともに幹葉表示を行う方法を学ぶ。

5 回目は、データ整理の方法として、グラフ表現の方法を説明する。まず、グラフの一般的な部分の名称（例えば、題目、単位、凡例など）を確認する。次に、構成比グラフと、構成比グラフの問題（構成比にすることで度数が分からなくなる）、そしてその解決方法（度数を記入する）について説明する。第 3 に、相対度数と累積相対

度数について理解する。第4に、度数分布表について説明して、棒グラフとヒストグラムの違いを理解する。最後に、ヒストグラムの作成方法を説明する。

6回目は、中心位置の統計量として平均値と中央値について理解する。まず、平均値と中央値の定義と計算方法を説明する。次に、平均値の性質として、(1)平均値からの偏差の和がゼロ、(2)仮の平均値を使った簡便な計算方法、そして(3)各区分の平均値と全体の平均値の関係の大まかに3つについて説明する。最後に、度数分布表が全体の分布状況を表すのに対して、平均値や中央値が中心位置を示す点を確認する。

7回目は、変化に関連した統計量を取り扱う。まず、変化を表現する用語を説明する。例えば、実数値の変化は増加や減少で表す。比率の変化で変化幅はポイントで表し、変化は上昇や低下で表す。変化幅の推移は、拡大や縮小で表す。これらの用語とともに、それぞれの計算方法も確認する。次に、変化率の分解について学ぶ。つまり、全データから求める変化率と、全データを構成する部分の変化率の関係について理解する。最後に、寄与度と寄与率について説明する。ある項目の変化率とその項目の構成比の積を寄与度と呼び、ある項目の寄与度を全体の変化率で除した答えを寄与率と呼ぶ。これらの概念と計算方法を学ぶ。

8回目は、中間テストであり、ここまでの学修事項をどの程度理解しているのかを学生自身がテストを通じて振り返られる機会としている。

3. 授業「経営統計論Ⅰ」の15回（半期）のうち後半7回の構成

ここでは引き続いて、9回の授業で取り扱う大まかな内容から順に確認する。9回目は、散らばりの統計量として、分散と標準偏差について学修する。分散へ向けて、まず中心位置からの偏差、絶対偏差、あるいは偏差平方を学ぶ。次に、絶対偏差の和を最小にするような中心位置の値として中央値、偏差平方の和を最小にする中心位置の値として平均値が対応することを確認する。その上で、偏差平方の和を合計した個数で均した値として分散を理解し、更に単位をデータの単位と同じ単位にした統計量として標準偏差を理解する。こうした概念的な理解をした上で、分散と標準偏差の計算方法について演習問題を解くことで習得する。

10、11回目は、標準偏差の活用である。9回目で学修した標準偏差は、概念が理解でき計算方法に慣れても、実際にどのように使えるのかがイメージしにくい。これは、散らばりという概念がイメージしにくいことと同意とも思える。そこで、散らばりがどのように使えるのかということで、チェビシェフの不等式、変動係数、そして標準化の3つで標準偏差の活用を理解する。チェビシェフの不等式は、データについて平均値と標準偏差という限られた情報しかない中で、どのような分布であったとしてもある区間に含まれる相対度数の下限を求められる。標準化は、データの中での観測値の相対的な位置をしれる。こうした利活用の方法を習得する。

12回目は、散らばりのグラフ表現を理解する。散らばりは、平均値に対して分散や標準偏差によりとらえられた。これに対して、中央値に対して四分位値や四分位範囲を求めることで散らばりをとらえる。更にこれをグラフ表現するときには、箱ひげ図が使われる。12回目の授業では、最終的に箱ひげ図の書き方について演習問題を

通じて理解していく。

13、14 回目は、2 変数の関連性ということで相関係数について理解する。2 変数の関連性として、共分散そして相関係数について説明する。2 変数のグラフ表現として、散布図を説明し、平均値からの偏差の積により 2 変数の関連性を表せることを説明し、共分散の計算方法を理解する。共分散の単位の問題から、標準化したものとして（ピアソンの積率）相関係数を説明する。相関係数の性質として、-1 から 1 までの値を取ることを、直線関係の強さを測るのであり関係性一般を測るのではないこと、そして相関関係と因果関係は異なること。これらを説明する。

15 回目は、これまで学んできたことの総復習として演習問題を自身の力で解く。問題数はそれほど多くないものの、解いてみることで学生自身が実際に問題を解けることを確認する。以上の 15 回の授業を通して、記述統計学の基本的な考え方や計算方法を習得できるように選択したテキスト（稲葉, 2012）にもとづいて授業している。では、各回の構成について節を改めて説明する。

4. 授業「経営統計論 I」の各授業回の構成

各回の標準的な授業の構成は、図 1 のような次の 3 ステップに分けられる。まず、予習課題の回収である。予習課題の一例は図 2 にある。この予習課題は、統計学の学修に必要な単語を中心にして、あらかじめインターネットや図書により単語の意味を調べてくる課題である。この課題の作成と提出が、出席票の代わりにとなる。つまり、出席とは着席していたことではなく、予習課題をこなした上で教員へ提出して、授業へ参加したこととなる。この課題は PDF 形式で久留米大学の e-learning システム上に置いてあり、受講生は自由にダウンロードできる仕組みになっている。

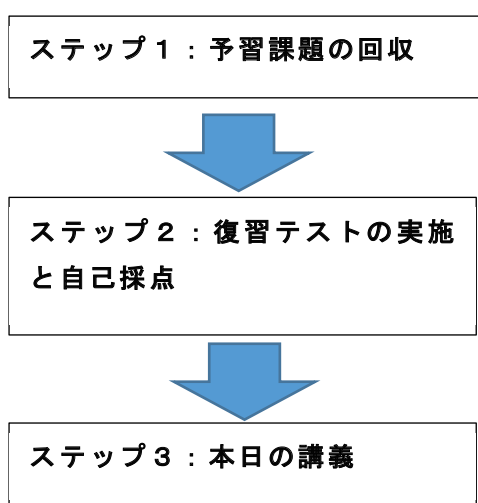


図 1. 標準的な授業の構成（出典、著者作成）

第 1 回 経営統計 I 予習課題

学籍番号：_____ 学年：_____

氏名：_____

※ 注意 ※

この課題の提出が出席票の代わりとなります。課題提出をしなければ、教室にいても出席とは評価されません。

テキストに出てくる単語のみを取り上げていますけれども、テキストの 1 単元分と対応していません。

課題：①以下の用語の意味を調べて、各用語のコロン（:）の後ろに書きなさい。②単語一語で答えてはいけません。③出典：以降に開けた文献名を書きなさい。Wikipedia と NAVER まとめを使ってはいけません。

(1) 質的データ：_____

出典：_____

(2) 量的データ：_____

出典：_____

(3) 名義尺度：_____

出典：_____

(4) 順序尺度：_____

出典：_____

(5) 間隔尺度：_____

出典：_____

図 2. 予習課題の一例（出典、著者作成課題を引用）

次に、復習テスト、自己採点、そして復習テスト自己管理票への記入である。復習テストは、前回授業内で解いた問題と全く同一の問題を小テスト形式で解かせることを指している。この復習テストにより、前回の学修内容を理解できているのか、各自が自分自身で確認できる仕組みになっている。理解を振り返られる仕組みという意味で、復習テストと呼んでいる。なお、この復習テストは、自己採点して、その結果を各自に手渡している復習テスト成績自己管理票に記入することで各自の責任で管理させている。この復習テスト成績自己管理票は、最終授業回に回収して成績へ反映させることをあらかじめ伝えてある。

以上のことが終了した後、授業内容へと入る。各回の授業は、記述統計の説明と演習問題から構成されている。このどちらもが基本的にテキスト（稲葉, 2012）にしたがって行われている。ここで演習問題が次の授業の復習テストに出題されることになる。その意味で、演習問題を中心とした授業となっている。なお授業で扱う内容は、既に 15 回の構成で示したように、記述統計の範囲である。

III. 考察

1. 学生の評価における信頼性と妥当性：現行シラバスの問題点

ここまでで現在著者が行っている経営統計論 I の授業内容を確認した。記述統計学は、推測統計学など他の統計学の範囲に比べて易しく理解しやすい。計算に必要とされる数学的知識は、加減乗除の四則演算程度であり、数学的に難しくて計算できないということも考えにくい。こうした特徴を念頭に置いて、本章では今後の統計学教育（とりわけ、経営意思決定に統計学を利用するという立場からの統計学教育）の方向性を探る。

まず、現行シラバスの問題点を確認する。経営統計論 I の受講生は統計学者ではなく、むしろ統計ユーザを目指している。この目指している目的と現行シラバスの内容は必ずしも一致していない。例えば、データが質的データや量的データに分類できたとして、経営意思決定にどう関わるのか。代表値として平均値や中央値を理解して、経営意思決定でどのように使うのか。現行シラバスから直接は導けない。

こうした意味において、現行のシラバスは具体的な経営意思決定の状況を仮定して、その状況下で統計学を用いて経営意思決定を行うような内容にはなっていない。この点は、今後再考しなければならない現行シラバスの問題点である。とはいえ、どのように再考すれば良いのか、以下でその手がかりを教育学者や教育の実践家の間で議論されているパフォーマンス評価に求める。

ここで、議論の出発点を学生の成績評価の信頼性と妥当性の問題に定める。教員が学生の成績評価を行う場合、その評価の信頼性と妥当性の 2 つの面について考えなければならない²。信頼性とは、誤差の入る割合の小ささであり、評価結果の安定性を示している³。この信頼性は、数値で測れる指標になっている⁴。これに対して、妥当性とは、評価したい対象をどれだけ本当に評価できているのかを示している⁵。

信頼性と妥当性という言葉が出てきたが、初見の方にとってはイメージしにくい概念だと思われる。そこで、信頼性と妥当性の違いを著者なりに、身近な例を用いて

直感的に示してみる。ここで仮にある人の体重を評価（測定）したいとする。このとき、この人の身長を評価（測定）したとする。身長計（身長測定の器具）を用いれば、身長は何度評価（測定）しても同じ値として評価（測定）でき、誤差の入る割合が小さく、評価（測定）結果は安定している。よって、身長は信頼性の高い指標となる。

だが、体重という評価（測定）したかった対象をきちんと評価でき（測れ）ているのかという点では疑問が残る。恐らくは身長と体重の間には相関関係がありそうだと想像できる。それでも、体重計を用いたときのように、直接的に評価（測定）したい対象を評価（測定）できているとはいえない。つまり、体重を評価（測定）したいときに、身長計で身長を評価（測定）する。これは、信頼性は高いものの、妥当性が低い評価（測定）方法である。このように考えられる。

さて、このように信頼性と妥当性について説明してくれば、議論の出発点を信頼性と妥当性に定めた理由が明らかになる。当初に示した現行シラバスに立ち返って説明してみる。現行シラバスは、信頼性の高い評価を目指している⁶。だが、妥当性の高いシラバスになっていない⁷。例えば、データが質的データかあるいは量的データかを区別させる問題を出し、正しく区別できていれば正解そうでなければ不正解と採点する。平均値を計算させて、正しければ正解そうでなければ不正解と採点する。

これらは信頼性の高い評価を可能にする。だが、現行シラバスは、経営統計論Ⅰが目指す（評価したい、あるいは測りたい）経営意思決定において統計学の知識を使うという、現実的な状況での知識の利用能力とでも呼ぶべき評価したい対象を評価できているとはいえない。つまり、妥当性が低いと思われるのである。

こうして説明してみれば、評価における信頼性と妥当性の問題が、本節の冒頭において述べた統計ユーザとして経営意思決定に統計学の知識を用いるという教育の方向性を探るのに適切な問いであると理解できるはずである。では引き続いて、教育学者や教育実践家がこの信頼性と妥当性の問題にどのように取り組んできたのか。とりわけ、妥当性の高い教育評価の実現に向けてどのように取り組んできたのか⁸。先行研究として、山口・石川(2012)、山口 (2013)、そして松下(2007)にもとづき手がかりを探る。

2. 妥当性のある評価：パフォーマンス課題、パフォーマンス、ルーブリック、そしてパフォーマンス評価

では、どのようにすれば教育評価において妥当性を確保できるのか。以下で、先行研究として、山口・石川(2012)、山口 (2013)、そして松下(2007)に依拠しつつ手がかりを探る。ここで真正の評価が注目される。真正の評価は、より現実的な状況設定の問題（リアルな課題）を解いてもらうことで、評価において信頼性を多少犠牲にしても妥当性を改善しようという評価である⁹。つまり、真正の評価での「真正」は、実験室実験のような限られた環境下で成り立つ法則ではなく、より現実的な文脈での課題により評価を行うという意味を込めて用いられている¹⁰。現実的な文脈での課題とは、別言すれば豊かな内包（ルール）と多くの外延（事例）を含む課題とされている¹¹。

こうしたより現実的な文脈での課題を通じて評価されるのは、課題解決に向けた

学生の思考、判断、あるいはスキルであり、これらは主として教員が観察可能な振る舞いとしてあらわれる¹²。現実的な文脈での課題を解くときの観察可能な振る舞いを「パフォーマンス」と呼び¹³、パフォーマンスを求める課題を「パフォーマンス課題」と呼び、そしてパフォーマンスを評価することを「パフォーマンス評価」と呼ぶ¹⁴。したがって、パフォーマンス評価は、現実的な文脈のもとで知識や技能を駆使して行う学生のパフォーマンスを直接的に評価することを意味する¹⁵。ここでパフォーマンス評価は、学生の実際のパフォーマンスを観察することから学力を評価するという考え方である¹⁶。

ここまですパフォーマンスというと、道具などを用いて実際に学生が体を動かしているある意味での運動を評価するように思えるかもしれない¹⁷。もちろんそうした実際に体を動かす方法でも良いが、別に統計学の知識を活用した教育や評価にも使える。それは、例えば次のような問題により可能となる¹⁸。現実世界の問題を統計学の知識で解決できる問題へと定式化する。定式化した問題を統計学の知識で解いて解を求める。最後に、その解を現実世界へと戻して現実世界の問題を解決する。このような方法である。

ここまですを整理してみる。まず、我々は経営意思決定に統計学を用いるような、より現実的な状況設定のもとでの経営統計学の教育を目指す必要性を述べた。そのような経営統計学の教育を目指していく手がかりとして、教育学者や実践家が教育評価における信頼性と妥当性の問題において、より妥当性の高い評価方法を模索する中で提案されている方法に注目した。それがより現実的な状況設定のもとでの学生の思考、判断、スキルといった直接観察可能な振る舞いをパフォーマンスと呼び、そのパフォーマンスを求めるようなパフォーマンス課題を通じて、パフォーマンス評価を行うという方法であった。このように整理できる。そして、このパフォーマンス評価で利用されている評価基準がループリックである¹⁹。

つまり、これまでのパフォーマンス、パフォーマンス課題、パフォーマンス評価、そしてループリックの相互関係を一文で表現し直すと次のようになる。パフォーマンス評価は、学力という直接観察できない能力をパフォーマンス課題により観察可能なパフォーマンスにし、そのパフォーマンスをループリックという評価基準に照らしながら解釈をしていく評価方法である²⁰。このようになる。では、これら概念を一枚の図に整理してみる。次の図3である。

まず、図3の中央に吹き出しがあり、「統計学の知識を経営意思決定に活用する能力」とある。これが学生の能力を表している。だが、この学生の能力は、教員から目で見て確認できない。そこで、教員は図の左端の「パフォーマンス課題」を学生に対して示し、その課題の成果として教員の目で見て確認できる「パフォーマンス」を獲得する。教員は、このパフォーマンスと教員が自身であらかじめ大枠で準備した評価基準の「ループリック」とを照らし合わせながら「パフォーマンス」に対する評価を確定させていく。ただし、ここで「パフォーマンス」と「ループリック」の間の矢印が双方向になっているところが、評価されるもの（パフォーマンス）と評価する基準（ループリック）との相互依存性を示しており重要である。ここまですの全体を図3では破線で囲み、この破線の中を「パフォーマンス評価」と呼ぶ。このように図示でき

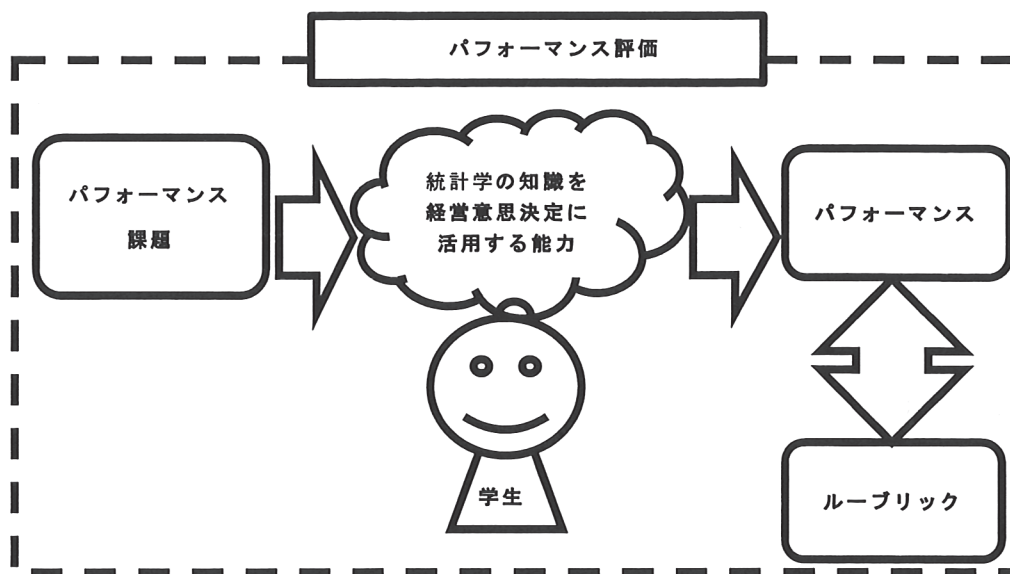


図 3. パフォーマンス評価をめぐる諸概念の関連図（出典、著者作成）

る。

3. パフォーマンス評価と客観テスト評価の違い

ここまででパフォーマンス評価について説明してきた。だが、パフォーマンス評価自体を取り上げていくら説明してみても、これまでの教育で行われてきた評価とどう違うのかという疑問がつきまとう。そこで、パフォーマンス評価と客観テストによる評価の違いを取り上げる。

ここで客観テストとは、紙と鉛筆により（別言すれば、ペーパーテストにより）ただ一つの正解に至るか否かを採点されるテストである。両者の違いは、(1)パフォーマンス課題と客観テスト問題の違い、そして(2)ルーブリック（パフォーマンス評価における評価基準）と客観テストの評価基準の違いの 2 つの観点で比較できる。この 2 つの点で比較することにより、パフォーマンス評価について一層明確にしてい

(1) パフォーマンス課題と客観テスト問題の違い

先行研究に依拠しつつ、パフォーマンス課題と客観テストの問題の違いをいくつか列挙してみる。まず、パフォーマンス評価は、問題を定式化し、その式を解き、解答を解釈し、コミュニケーションするという全プロセスを対象とする。パフォーマンス評価は、いわば総合的で複合的な知識の活用の評価を目指せる²¹。これに対して客観テスト問題は、こうしたパフォーマンス課題の対象とする全プロセスのうち一部分のみを対象としており、加えて結果のみを評価の対象とする

次に、パフォーマンス課題は、ペーパーテストだけでなく、物や道具を用いた問題解決プロセスを評価するものやグループで取り組ませるものもある。これに対して客観テストは、紙と鉛筆によりただ一つの正解に至るか否かを採点される。

更に、パフォーマンス評価が、時間を十分に与えて持てる力が十分に発揮されるよ

うにしたパワーテストの性格を持つ²²。これに対して客観テストは、ペーパーテストで、限られた時間で多くの問題を解くスピードテストの性格を持つ²³。ここまでに言及したものも含めて、松下(2007)はパフォーマンス評価と客観テストによる評価の違いを次の表2のような表に整理している。

表 2. パフォーマンス評価と客観テストによる評価の比較（出典：松下(2007、44 頁、表 5) を一部改変して使用）

	パフォーマンス評価	客観テストによる評価
評価手段	具体的活動、ペーパーテスト	ペーパーテスト
解答形式	自由記述式（ペーパーテストの場合）	択一式、簡単な記述式
問題数	少ない	多い
テストの性格	パワーテスト的性格	スピードテスト的性格
評価尺度	複合的な学力を、多次元的な尺度で、評価する（複数のレベル）	比較的限定された学力を、一元的な尺度で、評価する（正誤の二分法）
評価視点の設定時期	どんな学力が発揮されるかは、事後的に個人ごとに明らかになる（採点と同時並行で作成したルーブリックを用いる）	どんな学力を見る問題かをあらかじめ決めておく（採点コードを用いる）

このように比較してみると、パフォーマンス課題は優れており、客観テストは劣っていると誤解を生むかもしれない。しかし、そうではない。パフォーマンス評価は、学生の学力という直接観察できない評価対象を直接観察可能な評価対象（パフォーマンス）へと変える。このとき、観察可能なパフォーマンスへ変換しやすい学力もあれば変換しにくい学力もある²⁴。換言すれば、パフォーマンス課題でとらえにくい学力もある。また、客観テストは問題作成と採点の手間を省ける²⁵のに対して、パフォーマンス課題や評価は信頼性が低くなる危険性があり課題作成と評価に手間がかかる²⁶。こうした特徴も指摘されている。つまり、パフォーマンス課題は、万能薬ではない。

こうしたパフォーマンス課題のデメリットも考慮に入れた上で、より現実的にパフォーマンス課題を作成する。このように考えを進めるならば、ペーパーテストのような形式を活用したパフォーマンス課題の作成が考えられる。実際に、そうした試みも議論されている。つまり、パフォーマンス課題は客観テストと同じようなペーパーテストによっても行われている。そのときにはパフォーマンス課題と認められるような要件が示されている。その要件とは、思考プロセスの表現を求める、多様な表現方法が使える、真実味のある現実世界の場面を扱っている、その場面を定式化するプロセスを含む、複数の解法が取れるなどがあげられている²⁷。だが、こうした要件を満たす課題の作成は容易ではない。

ここまでの議論を前提として、今後統計ユーザとして経営意思決定に統計学を利用する能力を評価できるような課題の作成の手がかりとなりそうな具体例としては、パフォーマンス課題として評価を受けている問題がある。それは、どちらも経営統計学とは直接関係しないものの「平成 21 年度全国学力・学習状況調査の調査問題 小

学校国語 B²⁸」と「平成 19 年度全国学力・学習状況調査の調査問題 小学校算数 B 問題 5²⁹」である。これらは、パフォーマンス評価の具体例として示されている³⁰。したがって、今後、これらの問題の良い面を参照しながら、経営統計学において、統計ユーザとして経営意思決定に統計学を利用する能力を評価できるような課題の作成の一つの手がかりにできる。

(2) ルーブリック（パフォーマンス評価における評価基準）と客観テストの評価基準の違い

ルーブリックは、学習者の達成度を測る基準とされる³¹。このように説明すると、ルーブリックと客観テストの評価基準とはほぼ同じように見えてくる。しかし、むしろルーブリックと客観テストの評価基準は対比されて用いられる³²。そして、客観テストの問題点の解決の方法として提案されたのがルーブリックである³³。では、より詳細にルーブリックを理解する目的で、その作成手順について概説してみる。

ここでは学生を教員が評価するという状況で説明する。まず、学生が課題に対してどのようなパフォーマンスをするのかについて、教員はあらかじめ予想して可能な限り列挙しておく³⁴。次に、複数人の教員が学生の実際のパフォーマンスを段階に分けて採点する³⁵。この採点に当たり複数の教員がその段階と採点した理由をつきあわせて、合議の上で採点基準（ルーブリック）を練り上げていく³⁶。最後に、これらの作業をしつつ、各段階を最後まで埋めていく³⁷。

つまり、ルーブリックはあらかじめ予想して採点基準として準備しておく。だが、そのルーブリックは最終決定版ではなく、実際の採点において採点者間の合議を通じてルーブリックの改訂を繰り返しつつ採点を進めていく。ルーブリックは大まかに作っておき、採点と同時並行して評価基準を作っていく³⁸。

したがって、当初に想定していなかったようなパフォーマンスが出てきたとき、それをルーブリックに付け加えることも行う³⁹。ルーブリックは学生のパフォーマンスを予め予想して作成していくが、予想したパフォーマンスのうちひとつだけを使うものは少なく実際のパフォーマンスはもう少し複雑である⁴⁰。これは、学生の多様な学力の質を把握する方法とされている⁴¹。ルーブリックは、パフォーマンスの特徴に対応して得点を割り振る⁴²。

このように説明してみると、ルーブリックは教員の負担が重く⁴³、準備しなくても良いのではないかという疑問もわいてくる。だが、パフォーマンスを採点しようとすると、予めの構えとしてルーブリックの準備がないとパフォーマンスを適切に読み取れない⁴⁴。しかし、手間をかけて準備するにすれば、客観テストに比べて主観が入り込む余地を感じさせる。

これに対して、客観テストにも主観の入る余地があるとの反論もなされている。例えばどのような問題にするか、得られた得点をどう見るかといったところで、主観が入り込む⁴⁵。つまり、どのような評価方法を採用するにしても、評価者の主観が含まれる⁴⁶。パフォーマンス評価は、主観が見えやすいと理解できる⁴⁷。ただし、パフォーマンス評価に主観が含まれるとしても、恣意的あるいは独断的であってはならない⁴⁸。

そのための工夫として松下(2007、51頁)は次のような点を示している。「(a)ループリックによる採点基準の共有化(b)複数の採点者間でのモデレーション(採点結果が一致するよう調整する)(c)採点者のトレーニング(d)採点事例の蓄積・提供」つまり、一個人の主観的な評価ではなく、複数の教員による間主観的な評価と、評価基準や具体的事例のような評価の根拠を示せる必要がある⁴⁹。

パフォーマンス評価は、単に採点をして各学生の得点を出すと言うだけでなく、各学生のパフォーマンスを解釈することで、教員が学生の知識、理解、思考などを把握する助けとなる⁵⁰。学生にとっても、パフォーマンス課題にじっくりと取り組み、思考し、表現した成果を友達と比較するような機会を設ければ、自分と共通している点や異なる点を学び合うという学修経験を提供することもできる⁵¹。

ループリックの作成と採点をする、教員は学生の思考や表現方法、あるいはどの箇所で躓いているのかを理解できる⁵²。評価は単に得点をつけるという作業ではなく、教員が重視しているものは何かというメッセージを学生に伝えることでもある⁵³。パフォーマンス評価を指導に活かすには、こうしたパフォーマンス課題の作成からパフォーマンスを読み解くまでの作業を粘り強く実践していく根気が必要となる⁵⁴。そして、パフォーマンスを読み解くプロセスは、教員の学びについての鑑識眼を鍛えることにも繋がる⁵⁵。また、こうしたパフォーマンス課題の作成やループリックの作成という作業がそのまま教材研究にもなる⁵⁶。こうした作業の蓄積が、大学・学部独自の教育力になる⁵⁷。

ここまでの考察で、まだ不十分だが、パフォーマンス課題を経営統計学の文脈で作成し、そのパフォーマンス課題により現実的な文脈のもとで表れたパフォーマンスを、経営統計学の評価に使えるように作成されたループリックにより評価・採点する。こうした方向へ進むことにより、統計学の経営意思決定に活用する能力を教育・評価できるようになるのではないかという一つの手がかりを得た。

IV. 結論

1. まとめ

本論文の目的は、私立大学文系学部での著者自身の統計学教育の実践例を議論の出発点として例示した上で、将来の統計学教育(とりわけ、経営意思決定に統計学を利用するという立場からの統計学教育)の方向性を探ることであった。本論文を通じて、まず著者自身の授業として経営統計論Ⅰの目的、目標などを含めてシラバスを紹介した。

次に、この現行シラバスで行われている教育の問題点として、受講生が統計学者ではなく統計学を経営意思決定に利用したいと考えている統計ユーザであるとしたときに、その受講生の方向と現行シラバスの実現できている教育の方向が一致していない点を指摘した。したがって、今後、経営意思決定に統計学を利用するような授業を目指す具体的な教育方法を探らなければならない。

このような問題意識を明確にした上で、本論文では教育学者あるいは教育の実践家により示されてきたパフォーマンス評価という考え方に依拠した教育によりこれ

ら問題への解決の糸口がつかめるのではないかという 1 つの手がかりを示した。論文中にも記述したように、この方法が万能薬というわけではないが、現行のシラバスにより教えている統計学の知識ではなく、授業においても現実的な状況設定のもとで示されたパフォーマンス課題を解き、それをパフォーマンスという形で可視化し、教員があらかじめ作成したループリック（評価基準）とパフォーマンスとを行き来しつつ評価を決めていく。

こうした教育により、経営意思決定に統計が利用できるような能力を身につけさせあるいはそのような能力が身についているか否かを評価できるようになる。評価を見直すことは、単に採点を見直すことではなく、教育全体を見直すことになる。今後も引き続いて、こうした教育の可能性を探っていく必要がある。その 1 つの手がかりを示した。

2. 研究としての統計学教育

本論文は、著者自身の経営統計論 I の授業を出発点としつつ、今後の教育の方向性について論じた。こうした論文が、教育学者ならば兎も角も、マーケティング研究者により執筆されることは珍しいと言わざるを得ない。ならば、なぜ本論文は書かれたのか。もちろん既に目的で述べたように、データサイエンスが注目される中で学問分野として注目するだけでなく大学でどのように学修（あるいは教育）するのかを考える必要があるからである。だが、ここまでの議論を与件とするならば、もう 1 つの理由を示せる状況が整ったと言って良い。

それは、「Ⅲ 考察」において示した、学生の評価に関連する議論が、マーケティングにおける市場創造と深いところで繋がるからである。ここで学生の評価において示されたループリックという試みを思い返してみる。これは学生の評価において、あらかじめ定めておいた基準を外側から当てはめて学力の高低を得点として把握すること（いわゆる客観テスト）とは異なっていた。学問的には妥当性の問題として関連していた。

これは、別言すれば、測定の問題一般に見られることである。学生の学力を測定するだけでなく、消費者ニーズを測定する。このような場合にも、同様に消費者ニーズを測定するための質問項目をあらかじめ作成しておいて、その基準（あるいは測度）を外側から当てはめてニーズの高低を把握する。学生の学力測定に、同様の表現で当てはめ直してみれば次のようになる。学力を測定するためのテスト問題をあらかじめ作成しておいて、その基準（あるいは測度）を外側から当てはめて学力の高低を把握する。このように併記してみれば、対象は（教育の場合は）学生の学力と（マーケティングの場合は）消費者ニーズと異なっているものの、その深部において共通している大きな問題の枠組みのようなものを取り出せる。

この大きな問題の枠組みが、ここで測定の問題一般に見られる重要な問いである。そして、この問いに対して、我々が参照した教育学の先行研究は、大まかな基準を決めつつも実際に提出された答案とその大まかな基準とを行き来しながら、少しずつ基準それ自体を作り上げつつ採点するという試みを行っていた。これをループリックと呼んでいた。

こうした教育学での試みはマーケティングにおいても同様に、消費者ニーズが測定より先に存在しているのではないという認識と共通する。つまり、著者のこれまでの研究をご存知の読者に向けて言えば、評価（主成分）軸それ自体の変化をもって市場創造と考えるという研究のスタンスと評価軸を与件としないという点で共通している。したがって、マーケティングの研究者が本論文のような一見すると教育学の専門家を対象としたような論文に関心を持ち執筆するのも理解できるはずである。

参考文献

- 稲葉由之(2012). 『プレステップ 統計学Ⅰ 記述統計学』, 弘文堂.
- 加藤淳一(2017). 「経営統計論Ⅱ」, 『平成 29 年度学修シラバスブック 〈商学部〉』, 久留米大学, 151 頁.
- 松下佳代(2007). 『パフォーマンス評価ー子供の思考と表現を評価するー』, 日本標準.
- 山口陽弘(2002). 『試験にでる心理学: 心理測定・統計編』, 北大路書房.
- 山口陽弘・石川克博(2012). 「教育評価の理論と実践ー真正の評価をめざしてー」, 『群馬大学教育実践研究』, 第 29 号, 187-200 頁.
- 山口陽弘 (2013). 「教育評価におけるルーブリック作成のためのいくつかのヒントの提案ーパフォーマンス評価とポートフォリオ評価に着目してー」, 『群馬大学教育学部紀要 人文・社会科学編』, 第 62 巻, 157-168 頁.

¹ 滋賀大学データサイエンス学部については次の URL で紹介されている。

<https://www.ds.shiga-u.ac.jp/> 最終アクセス日: 2017 年 10 月 27 日

² 山口・石川(2012, 192 頁)は、「その統計学的知識のもっとも重要な核が、信頼性と妥当性であると考えている（信頼性と妥当性関しての詳細は山口(2002)を参照）。」と述べている。

³ 山口・石川(2012, 192 頁)は、「信頼性とは、教育評価に即して言えば、評価結果が安定している度合いである。誤差が評価に入ってくる割合が少ないほど信頼性が高いと言える。」と言及している。

⁴ 山口・石川(2012, 192 頁)は、「よく用いられるものが再検査信頼性やクロンバックの α 係数である。前者は、テスト（評価）を同一人に二回繰り返しても同じ値を示すとき、そのテストは再検査信頼性が高いと言われる。後者は、内的整合性、一貫性とも言われるもので、テスト項目間で整合性がとれていて、一貫性がある場合に、 α 係数は高くなり、優れた項目であると言われる。信頼性係数として、何をどのような目的で使うのかによるが、目安として最低でも 0.7 以上、できれば 0.8 以上の数値を得る事が望ましいと言われている。」と述べている。

⁵ 山口・石川(2012, 192 頁)は、「一方、妥当性とは、真実性とも言われるもので、評価したい対象を、どれほど本当に評価できているかということを示す概念である。」と述べている。

⁶ 松下(2007, 50 頁)は、「客観テスト」の「客観」とは、採点の信頼性が高いということです。」と指摘している。

⁷ 山口 (2013, 163 頁)は、「この客観テストは「真正の評価」という目的からすると、しばしば否定的なニュアンスで扱われるが、長所もちろん、存在することはここでも再確認しておきたい。まず、第一に、客観テストは信頼性が極めて高いという長所がある。信頼性とは評価結果の安定性を意味する。このこと自体は推奨されるべき立派な長所である。注意すべきなのは、妥当性も高い、すなわち、当該単元の学習者の達成度がすべてここに反映されていると即断することである。」と指摘している。

⁸ 山口・石川(2012, 193 頁)は、「したがって、心理学者もどちらかといえばいかにして信頼性を上昇させるかという議論をしがちであるようである。しかし、心理学者も、もちろん信頼性よりも妥当性が重要であるという認識はしている。」と言及している。

⁹ 山口・石川(2012, 193 頁)は「では、教育評価という枠組みの中で、妥当性を上昇させるためにどうすればよいのだろうか。ここで再び「真正の評価」という概念が有効になってくる。信頼性を多少犠牲にしても、より日常的なリアルな文脈で、リアルな課題を解いてもらうことで妥当性を確

- 保すること、その上で評価するというのが真正の評価の意図であった。これは妥当性を向上させる際にも、そのまま使える考え方であろう。教育評価をよりリアルであるような文脈や課題にすることが、妥当性向上のためにもっとも直接的な方法である。」と指摘している。更に、松下(2007、50 頁)は、「パフォーマンス評価は、正誤問題や多肢選択問題のような客観テストに比べると、「信頼性」が低くなります。」と述べている。
- ¹⁰ 山口・石川(2012、188 頁)は、「すなわち、実験室的条件で得られる心理学的法則は、非常に特殊な状況下での法則であり、現実にはありえない。したがって、できるだけ日常文脈に即した、生態学的に妥当な状況で心理学実験は行われ、そこでの法則こそが意味あるものであるという主張である。以上をまとめれば、教育評価を「真正な」ものにするためには、より日常文脈に即した方法でリアルな課題に取り組ませる必要があるということである。」と述べている。
- ¹¹ 山口 (2013、164 頁)は、「この「真正の課題」とは、日常的なものとも繋がるようなリアルな課題ということである。内包（いわばその概念を獲得するための性質）が貧弱であったり、外延（概念の具体例である事例）がわずかしき存在しないような課題は、リアルな課題とは言えない。その概念の内包（性質、ルール）は豊かであり、多くの外延（事例）が含まれるような課題、つまり、日常と結びついているような概念を獲得するための課題が、「真正の課題」となっていくのである。」と言及している。
- ¹² 山口 (2013、160 頁)は、「しかし、よりリアルな課題であるような高次の思考・判断、スキルなど（＝パフォーマンス）の評価は、量的な形では難しい。したがって、そのために考え出された質的な、主として行動的な面で、第三者にも観察可能な行動指標上の特徴をもとにして作成されるのである。」と述べている。
- ¹³ 山口 (2013、163 頁)は、「パフォーマンス評価やパフォーマンス課題で言うところの「パフォーマンス」とは、現実に近いような場面を想定した「リアルな」課題を解くときの「振る舞い」すべてを総称するものである。」と述べている。
- ¹⁴ 山口 (2013、163 頁)は、「つまり「真正の課題」＝「リアルな課題」＝望ましい「パフォーマンス課題」ということになり、これらを実評価する方法が「パフォーマンス評価」なのである。」と指摘している。
- ¹⁵ 松下(2007、6 頁)は、「つまり、パフォーマンス評価とは、「ある特定の文脈のもとで、様々な知識や技能などを用いて行われる人のふるまいや作品を、直接的に評価する方法」のことです。」と述べている。
- ¹⁶ 松下(2007、7 頁)は、「これと同じように、実際に何かをやらせてみて（パフォーマンスさせてみて）、それによって直接的に学力を評価しようというのが、パフォーマンス評価なのです。」と指摘している。
- ¹⁷ 松下(2007、15 頁)は、「算数・数学のパフォーマンス課題でも、ハンズオン型（手で扱えるような道具を使って具体的な活動に取り組ませるもの）の課題が使われることがあります。」と指摘している。
- ¹⁸ 松下(2007、11 頁)は、「この図式では、数学（特に学校教育において学ぶべき数学）を、①直接には解決困難な現実世界の問題を数学の問題に「定式化」し、②その問題を数学的な「技法」で解いて解を求め、③この解を再び現実世界に「解釈」してもとの問題の解決とする、という活動ととらえています。」と述べている。
- ¹⁹ 山口 (2013、160 頁)は、「つまり、児童生徒が学習した結果、当該単元を代表するようなリアルな課題を解決する際に、実際に行える振る舞い（＝パフォーマンス、次章で再度詳述）のレベルの目安を、いくつかの段階に分けて記述して、学習の達成度を判断する基準を示すために作成されたものである。」と言及している。
- ²⁰ 松下(2007、10 頁)は、「学力をパフォーマンスのかたちにして見えるようにすることを「可視化」、パフォーマンスからその背後にある学力を推論することを「解釈」と呼ぶことにしましょう。そうすると、パフォーマンス評価とは、「パフォーマンス課題」によって学力をパフォーマンスへと可視化し、「ループリック」などを使うことによってパフォーマンスから学力を解釈する評価法だということです（図 2 参照）。」と指摘している。更に、松下(2007、23 頁)は、「このような課題によって「可視化」されたパフォーマンスを、ループリックを使って「解釈」し、評価していきます。」と述べている。
- ²¹ 松下(2007、4 頁)は、「なかでも、一般的な標準テストでは測りにくい質の学力（思考プロセスや算数・数学的なコミュニケーション能力など）も含めて算数・数学の学力を多面的・総合的に把握することを、主な目的としました。」と言及している。
- ²² 松下(2007、44 頁)は、「また、時間はたっぷり与えて、もてる力が十分発揮されるようなパワーテストの性格をもたせました。」と述べている。
- ²³ 松下(2007、43 頁)は、「パフォーマンス評価では、問題を数学的に定式化し、解決し、解釈するというプロセス全体を自分で展開し、それを表現できるかどうかをみるのに対し、ふつうのテストでは、プロセスの一部、しかもその結果のみをみます。ふつうのテストは、もっぱらペーパーテストで、問題数が多く、それを限られた時間内に解かせるスピードテストの性格をもっています。それに対して、パフォーマンス評価には、ペーパーテストだけでなく、具体的な物や道具を与えてそれを使った問題解決のプロセスをみるものや、グループで取り組ませるものなどもあります。」と指摘している。
- ²⁴ 松下(2007、9 頁)は、「つまり、(a)「見える」のはパフォーマンスであって、学力そのものはパフ

- パフォーマンスから推論することでしか把握できない(b)しかし、「見えやすい学力」と「見えにくい学力」はあって、「見えやすさ」の程度は、学力とパフォーマンスを結びつけるのがどれくらい容易かによって決まるということが出来ます(図1参照。)」と述べている。
- 25 山口 (2013, 164 頁)は、「第三の長所は、教師の問題作成、採点などの手間を省くことができる点である。」と指摘している。
- 26 山口 (2013, 164 頁)は、「つまり、パフォーマンス評価の最大の問題であり、欠点であるのが、その課題の作成および採点に非常に手間がかかる点である。また、「ルーブリック」が然るべき手続きに基づいていない場合に、その信頼性(評価の安定性)が低い場合もしばしば生じる。パフォーマンス課題の作成過程においては、より「リアル」であるような場面を想定して、しかも当該単元の中で、中心的な思考や判断や表現力などを評価するのに適した課題を作成する必要がある。」と述べている。
- 27 松下(2007, 17 頁)は、「ただし、できるだけパフォーマンス課題の要件を満たすように次のような特徴をもたせています。(a)思考のプロセスを表現することを要求する(b)多様な表現方法(式、言葉、図、絵など)が使える(c)真実味のある現実世界の場面を扱っていて、そこから数学化するプロセスを含んでいる(d)複数の解法がとれる」と言及している。
- 28 http://www.nier.go.jp/09chousa/09mondai_shou_kokugo_b.pdf 最終閲覧日: 2017 年 10 月 27 日
- 29 http://www.nier.go.jp/tyousa/07mondai_shou_sansuu_b.pdf 最終閲覧日: 2017 年 10 月 27 日
- 30 山口・石川(2012, 196-198 頁)において、真正の評価にもとづくパフォーマンス評価の例として示されている。
- 31 山口 (2013, 157 頁)は、「なるだけ簡潔に述べれば、ルーブリックとは、「学習者の達成度を示すための基準」を意味する。」と述べている。更に、山口 (2013, 160 頁)は、「ルーブリック(rubric)とは、いくつかの段階に分けて教育上の達成度の目安を記述して、学習者の達成度を判断する基準を示したものである。」と指摘している。
- 32 山口 (2013, 160 頁)は、「このルーブリックと対比されるのが、前章でも触れた客観テスト(いわゆる○×式のテスト)の量的な結果(70 点とか 60%解答したとか)である。」と指摘している。
- 33 山口 (2013, 163 頁)は、「このように、そもそもパフォーマンス評価は、客観テストとの対比から、その問題点、限界点を克服するために提案された評価方法である。客観テストとは、いわゆる紙と鉛筆で測定され、しかも模範的な解答が一つに決まっており、その解答を採点するようなテストである。」と述べている。
- 34 山口 (2013, 161 頁)は、「ルーブリックの作成方針は次のような手順に基づくものである。まず、第一に、事前に予想される児童生徒の様々な振る舞い方(問題、課題への解法)を、可能な限りリストアップしておくことである。」と述べている。
- 35 山口 (2013, 161 頁)は、「第二に、可能であれば複数の教師(できれば 3 人くらい)で、そうしてリストアップされた解答パターンや振る舞い方を段階に分けて採点する。」と言及している。
- 36 山口 (2013, 161 頁)は、「その際に、複数の教師の間で、その段階にした理由をつき合わせて、合議の上で練り上げていく。」と述べている。
- 37 山口 (2013, 161 頁)は、「最後に、これらの作業をしながら、ルーブリックの各段階のすべての段階を埋めていく。」と指摘している。
- 38 松下(2007, 27 頁)は、「ルーブリックは、採点に入る前にあらかじめ大まかな素案を作っておきますが、具体的な中身は、採点と同時並行で書き込んでいきます。つまり、パフォーマンス評価には、〈ルーブリック作成＝評価基準作り〉というプロセスが含まれるのです。この点が、すでにできあがった採点基準にそって採点していくフィギュアスケートの場合と大きく異なるところです。」と述べている。更に、松下(2007, 44 頁)は、「他方、パフォーマンス評価では、ルーブリックを採点と同時並行で作成して評価します。これによって、解答に表れた複合的な学力を、多次元的な尺度で評価するわけです。どんな学力が発揮されているかは、事後的に(テストが実施された後に)、個人ごとに明らかになります。こうして個々の子どもの個性的な学力の質を把握することが可能になるのです。」と言及している。
- 39 山口 (2013, 161 頁)は、「新しい解答パターンが生まれた場合、ルーブリックに追加する作業を継続していく(cf.松下, 2007)。」と指摘している。
- 40 松下(2007, 29 頁)は、「表 1 に事前に予想した解法をあげましたが、実際に採点してみると、予想した解法のどれか一つだけを使ったものは意外に少なく、もう少し複雑な様相を呈していることがわかってきました。」と指摘している。
- 41 松下(2007, 38 頁)は、「パフォーマンス課題によって学力は観察可能なパフォーマンス(解答)へと可視化され、そのパフォーマンスが解法のタイプとルーブリックというツールを使って解釈・評価されます。こうしたプロセスをたどって、個々の子どもの多様な学力の質を把握できるようになるのです。」と指摘している。
- 42 松下(2007, 23 頁)は、「どんな特徴が見られれば、どんな数値(得点)をわりあてるかを述べたものが、ルーブリックです。ルーブリックには、ある教科や領域で共通する「一般評価基準」と、それをもとに課題ごとに作っていく「課題別評価基準」があります。」と指摘している。
- 43 山口 (2013, 164 頁)は、「また、その課題を評価するために「ルーブリック」を作成する必要がある、その上で採点していく必要がある。これらの労力は、教師にとっては現実的に負担が重いことである。」と述べている。

-
- 44 松下(2007, 30 頁)は、「この子は、どの解法を、どんなふうに使っているのか」。その構えと準備がないと、自由記述式の問題の解答はちゃんと読み取れません。」と述べている。
- 45 松下(2007, 50 頁)は、「しかし、客観テストであっても、どんな問題にするかとか、得られた点数をどうみるか、といったところには、必ず主観（価値判断）が入り込んできます。」と言及している。
- 46 松下(2007, 50 頁)は、「つまり、どんなかたちの評価であっても、そこには必ず評価者の主観が含まれるものなのです。」と言及している。
- 47 松下(2007, 50 頁)は、「このように考えれば、パフォーマンス評価の場合は、評価につきまとう主観性が、見えやすくなっているにすぎないということもできます。」と言及している。
- 48 松下(2007, 11 頁)は、「パフォーマンス課題は、評価したいと思う学力ができるだけ直接的にパフォーマンスとして表れるものにする必要があります。また、パフォーマンスから学力への解釈は主観的な性格を免れませんが、主観的であっても恣意的・独断的にならないようにしなければなりません。そのためにルーブリックなどを使うのです。」と言及している。更に、松下(2007, 51 頁)は、「もちろん、パフォーマンス評価では、主観的ではあっても恣意的・独断的にならないような工夫がなされています。これまでも述べてきましたが、あらためてまとめておくと、次のような工夫です。」と述べている。
- 49 松下(2007, 51 頁)は、「つまり、複数の評価者による評価結果のつきあわせ・交渉によって、複数の評価者の主観を交えた「間主観性」をもたせるようにする、採点基準や採点事例によって評価の根拠を示せるようにする、評価者自身の力量を高めるということです。」と述べている。
- 50 松下(2007, 48 頁)は、「しかし、パフォーマンス課題の開発は教材研究そのものですし、個々の解答の解釈は子どもたちの知識・理解や思考などの把握を助けます。」と述べている。
- 51 松下(2007, 65 頁)は、「一方、子どもたちにとっても、20 分間、1 問の課題についてじっくり思考し表現すること、書いたものを通じて友だちと自分の違いや共通性に気づき学びあっていくことは、意味のある学習経験であったと思います。」と指摘している。
- 52 松下(2007, 59 頁)は、「ルーブリック作成と採点のなかで、評価する側は、子どもたちがどんなふうに考え、表現しているか、どこでつまづいているかを知ることができます。」と述べている。
- 53 松下(2007, 64 頁)は、「どのような評価法を用いるかは、教師が本当に重視しているものは何であるかを、暗黙のうちに子どもたちに伝えます。評価とはそのようなメッセージ性をもつものなのです。」と言及している。
- 54 松下(2007, 64 頁)は、「パフォーマンス評価を指導に生かすには、パフォーマンス課題を作り、課題によって可視化されたパフォーマンスから学力を読み解くという作業を粘り強く行うことが欠かせません。パフォーマンス評価に携わる人間には、この作業に身を置く覚悟が求められます。」と指摘している。
- 55 松下(2007, 64-65 頁)は、「解答（パフォーマンス）から、それぞれの子どもの個性的な思考や表現を読み解き、評価していくには、教師の側に「鑑識眼」が要求されます。逆にいえば、パフォーマンス評価は、子どもの学びについての鑑識眼を鍛えていく機会を提供してくれます。」と述べている。
- 56 松下(2007, 65 頁)は、「また、パフォーマンス課題やルーブリック作成は、そのまま教材研究の場になります。」と述べている。
- 57 松下(2007, 65 頁)は、「パフォーマンス課題やルーブリック、採点事例を蓄積していけば、時間と労力は次第に軽減されていくはずで、そしてそれらの蓄積は、学校独自の貴重な財産になることでしょう。」と言及している。